

# 分析型数据库

快速入门

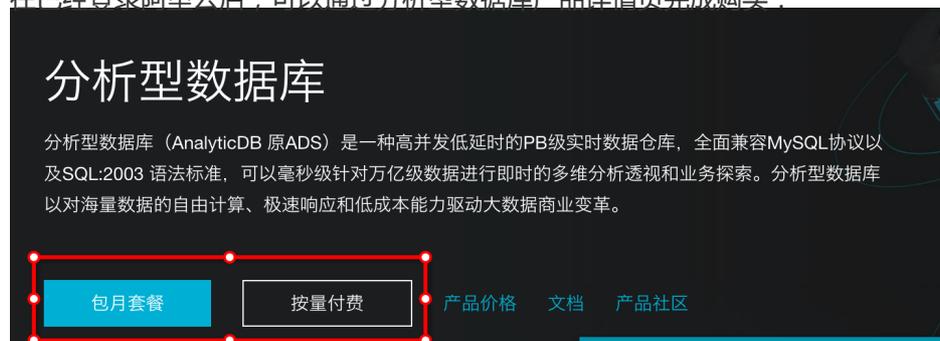
# 快速入门

## 1.1 开通分析型数据库服务

在公共云上，满足开通条件的用户：

1. 账户有100块余额
2. 已经实名认证通过

在已经登录阿里云后，可以通过分析型数据库产品详情页完成购买：



- 包月套餐：称为预付费，适合长期需求，价格比按量付费更实惠，且购买时长越长，折扣越多。
- 按量付费：称为后付费，按照使用的时间来付费，即每小时生成一个收费订单，并按照生成订单时的实例规格从阿里云账号扣费。

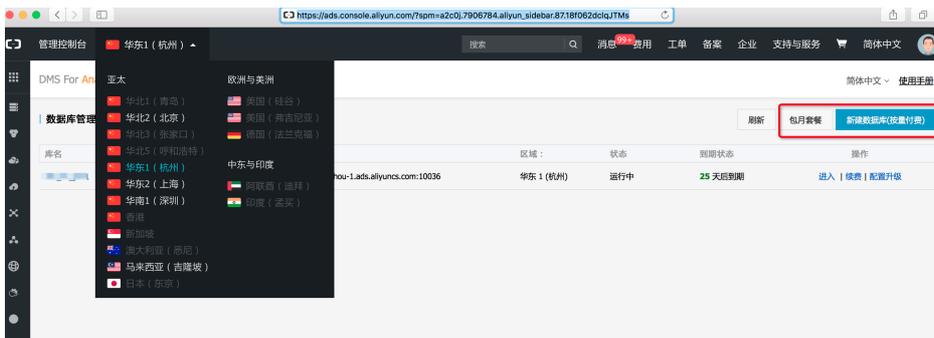
更详细的计量计费策略见[https://help.aliyun.com/document\\_detail/26434.html](https://help.aliyun.com/document_detail/26434.html)

## 1.2 创建数据库

登录阿里云官网，通过阿里云控制台首页的相关按钮，或直接访问<https://ads.console.aliyun.com> 来登陆DMS for AnalyticDB管理控制台创建数据库。

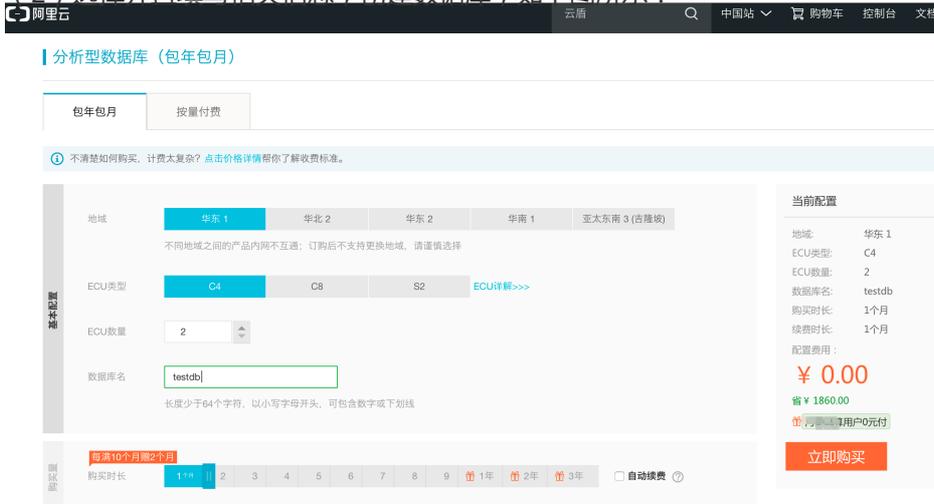
### 1. 通过DMS 创建数据库

(1) 通过“新建数据库”按钮，进入到创建数据库页面，如下图所示：



特别说明：上图中显示了当前ADS开通服务的region，如果当前相关区域还没有开通ADS服务，则相关区域呈灰色状态。

(2) 选择并且填写相关信息，创建数据库，如下图所示：



关于数据库的创建，有三点

特别说明：

首先，创建数据库时，需要填写数据库名，**注意这个数据库名称需要在分析型数据库全部集群上全局唯一**；

其次，选择分析型数据库的Region所在地，如杭州、北京等。

然后，关于ECU的选择。分析型数据库以ECU（弹性计算单元）作为资源单位。ECU拥有多种型号，每种型号的ECU，标识着不同的内存大小、磁盘空间大小。用户在创建数据库时根据自身需求选择ECU型号，以及初始的ECU数量（必须是偶数个，至少两个）。ECU数量可以在使用中随时调整（扩容/缩容），关于ECU的详细信息，详见 2.4节 ECU详解。

最后，填好所有选项后，点击创建数据库，几分钟后控制台<https://ads.console.aliyun.com> 会显示创新成功后的数据库地址。

需要注意的是，任何人都可以访问分析型数据库的控制台，但是没有开通分析型数据库服务的客户无法使用DMS中的创建数据库（按量付费）的功能（会提示您不是ADS用户）。作为数据库的owner或被授权使用某数据库的用户，以及被主账号在RAM中收于操作分析型数据库权限的用户，可以在控制台中看到数据库列表并进行管理。

### (3) 登录控制台。

开通分析型数据库服务并且资源分配好后，就可以登录DMS for analyticDB控制台（如下图所示）进行日常操作了，只需通过阿里云控制台首页的相关按钮，或访问<https://ads.console.aliyun.com> 即可。



最后，总结在创建数据库中需要注意的问题如下：

数据库名称需要在分析型数据库全部集群上全局唯一；  
ECU型号在DB创建后可以升降配进行调整、ECU数量可以在使用中随时扩容/缩容；且ECU数量必须是偶数个、至少两个；  
在创建数据库过程中，可能会遇到库存不足导致创建失败的问题，请及时提交工单。

## 1.3 使用DMS创建和管理表

前文中，假设我们已经创建了一个分析型数据库的数据库ads\_demo，本章节，将通过DMS控制台界面在ads\_demo数据库中创建一个表组和一个表。

### 操作步骤

在DMS中，进入ads\_demo数据库，右击左侧表组对象，选择新建表组，弹出新建表组对话框，弹出如下图所示内容，我们填写表组名为test\_group，其余参数先暂时使用默认值。



单击确定，建立好表组后，右击表组，选择新建表。在这里，我们根据测试数据的情况，建立一张有五列和一级分区的实时表。如下图所示：



表组名：选择前面创建的test\_group。

表名：表名称，这里填表名为qsdemo\_load\_data。

是否维度表：本例要创建普通表，所以不勾选。

更新方式：选择批量更新，即表数据来源是通过其他系统批量导入。（如果表的数据来源来自于直接的insert插入，那么此项选择实时更新）。

一级分区列：user\_id。在分析型数据库中，数据根据分区列进行分布式的存储和计算。

分区方式：进行hash分区，分区数为40，（一般情况下，每个分区数据量不超过3000万条为宜，分区数不能超过256，大于256提工单处理）。

列名：根据实际情况填写，如果是从MaxCompute等系统批量导入数据目前需要和源头数据表的字段名称一致。

索引方式：分析型数据库拥有强大的自动索引功能，在创建表时通常无需关心一个列的索引情况，分析型数据库会根据实际数据的分布情况来自动进行索引。所以这里我们先不调整列的索引设置。

单击保存，弹出实际的建表DDL供校验。创建表完毕后，右击已有的表可以进行编辑。

有关表和列的详细说明，请参见：创建和修改表。

## 1.4 导入数据

## 1.5 导入数据

分析型数据库支持多种接入数据的方式：

- 通过Kettle等ETL工具将本地文件写入分析型数据库。
- 通过阿里云数据传输从阿里云RDS中实时同步数据变更。（见《用户指南》相关使用阿里云数据传输实时同步RDS的数据章节—仅公共云可用）
- 建立批量导入表从阿里云MaxCompute（原名ODPS）大批量的导入数据。
- 通过阿里云数据集成（专有云中通过大数据开发套件的‘数据同步’）任务导入数据。（详见《用户指南》数据的导入章节）
- 直接将数据通过insert/delete SQL写入实时表。（详见《用户指南》insert章节）

### 批量导入

批量导入适合每天1-2次且每次同步大量批数据（每次200w条以上）。

特点：小规模数据量的同步速度相对很慢；大批量的数据同步速度相对较快，但是整体任务延迟长。

### 实时写入

实时写入适合数据写入后比较短时间内（例如1-2分钟）就需要查询的场景。

特点：每次写入的数据相对少（一条到几十条），每秒钟写入一般在千条到万条的数据量。（如果需要非常高的写入速度，例如达到百万条每秒，需要相对大量的ECU资源和优化的写入方法）。

在这里，作为测试使用，我们将介绍如何通过控制台界面进行数据批量导入和实时写入。

## 通过控制台批量导入数据

我们设定这里通过控制台界面进行MaxCompute（原ODPS）数据导入，AnalyticDB目标表即为前面创建的qsdemo\_load\_data表。

### 前提条件

AnalyticDB目标表qsdemo\_load\_data创建时更新方式属性为“批量更新”；

### 操作步骤

步骤1：数据准备。

此处我们需要把MaxCompute的表导入到分析型数据库，因此需要先在MaxCompute中准备好表和数据。

MaxCompute表建表语句：

```
use projecta ; --在MaxCompute的某个project中创建表
CREATE TABLE
demo_load_data (
user_id bigint ,
amt bigint ,
num bigint ,
cat_id bigint ,
thedata bigint
);
```

本地创建demo\_load\_data.txt文件，输入数据如下，列分割符为英文逗号，回车换行。

```
12345,80,900,1555,20140101
12346,81,901,1556,20140101
12347,82,902,1557,20140101
12348,83,903,1558,20140101
12349,84,904,1559,20140101
12350,85,905,1560,20140101
12351,86,906,1561,20140101
12352,87,907,1562,20140101
12353,88,908,1563,20140101
12354,89,909,1564,20140101
12355,90,910,1565,20140101
12356,91,911,1566,20140101
12357,92,912,1567,20140101
12358,93,913,1568,20140101
12359,94,914,1569,20140101
12360,95,915,1570,20140101
12361,96,916,1571,20140101
12362,97,917,1572,20140101
12363,98,918,1573,20140101
12364,99,919,1574,20140101
```

MaxCompute数据上传，若用tunnel上传命令如下：

```
use projecta;  
tunnel upload demo_load_data.txt demo_load_data;--把本地数据文件demo_load_data.txt
```

步骤2:在操作导入数据之前，需要对数据的来源表进行授权。

此处数据的来源表在MaxCompute上，所以需要在MaxCompute上对云账号授予describe和select权限。

各个专有云授权的云账号名参照专有云的相关配置文档，一般为test1000000009@aliyun.com

公共云上需要给garuda\_build@aliyun.com以及garuda\_data@aliyun.com账号授权。

授权命令：

```
USE projecta ; --表所属ODPS project  
ADD USER ALIYUN$xxxx@aliyun.com;--输入正确的云账号  
GRANT Describe,Select ON TABLE table_name TO USER ALIYUN$xxxx@aliyun.com;--输入需要赋权的表和正确的云账号
```

步骤3:进入DMS for AnalyticDB页面，选择菜单栏上的导入按钮，弹出导入对话框。这里我们的数据源表在阿里云MaxCompute（原ODPS）上。因此数据导入路径按照

“odps://project\_name/table\_name/partition\_spec”的格式来填写。目标表选前面创建的表qsdemo\_load\_data。填写完毕后，如下图所示，点击“确定”按钮。

关于分区信息，在仅有一级Hash分区的情况下DMS会自动识别并填写。

步骤4:接下来页面会显示导入状态一览，分析型数据库会对导入任务进行调度，根据当前系统繁忙情况和待导入数据的大小和结构不同，二十分钟至数个小时内数据导入会结束。

步骤5:数据导入结束后，可以在左侧栏对表进行右键“SQL操作”查看表的数据。

## 通过控制台实时写入数据

创建表时表属性“更新方式”选择“实时更新”，则可以在建表后直接在SQL窗口中编写SQL语句实时写入数据。

### 操作步骤

步骤1:创建表my\_first\_table，参考前面《使用DMS创建表组和表》章节，表属性“更新方式”选择“实时更新”。

步骤2:进入控制台的SQL窗口，输入命令：

```
insert into my_first_table (user_id,amt,num,cat_id,theodate) values (12345, 80, 900, 1555, 20140101);
```

步骤3:执行命令。

步骤4:命令执行成功后，可以在左侧栏对表进行右键“SQL操作”查看表的数据。

注意：

- 实时更新的表刚刚建立后，会有一分钟左右的准备时间，这时候写入的数据需要在准备完成后才能查询，否则查询会报错。
- 在准备完成后，实时进行insert/delete数据变更，一般需要等待一分钟后才可查询。
- 分析型数据库进行实时插入和删除时，不支持事务，并且仅遵循最终一致性的设计，所以分析型数据库并不能作为OLTP系统使用。

其他的数据导入方式请看《用户指南》相关数据导入章节。

## 1.5 在应用中连接和使用分析型数

首次成功导入数据到分析型数据库后，我们便希望我们的应用系统能够连接到分析型数据库来进行数据查询。分析型数据库可以通过任何支持 5.1.x 5.4.x 5.6.x协议的客户端进行连接。连接所使用的域名和端口号可以在iDB Cloud的右上角进行查看。连接使用的用户名和密码为用户在阿里云的Access Key，可以在[https://i.aliyun.com/access\\_key/](https://i.aliyun.com/access_key/) 查看和管理。其中Access Key ID为用户名，Access Key Secret为密码（分析型数据库承诺不会保存用户的Access Key信息）。

若需要使用阿里云访问控制(RAM)子账号连接分析型数据库，请参见使用RAM子账号访问分析型数。

### 在PHP中连接分析型数据库

在PHP环境下，假设我们已经安装好了php-mysql 5.1.x模块（Windows下为php\_MySQL.dll），那么我们新建一个ads\_conn.php，内容如下：

```
$ads_server_name="mydbname-xxxx.ads-cn-hangzhou-1.aliyuncs.com"; //数据库的连接url，请在控制台中的连接信息中获取
$ads_username="my_access_key_id"; // 连接数据库用户名
$ads_password="my_access_key_secret"; // 连接数据库密码
$ads_database="my_ads_db"; // 数据库的名字
$ads_port=3003; //数据库的端口号，请在控制台中的连接信息中获取
// 连接到数据库
$ads_conn=mysqli_connect($ads_server_name, $ads_username, $ads_password, $ads_database, $ads_port);
```

执行查询时，可以使用：

```
$strsql="SELECT user_id FROM my_ads_db.my_first_table limit 20;"; $result=mysqli_query($ads_conn, $strsql);

while($row = mysqli_fetch_array($result)) {
    echo $row["user_id"]; //user_id为列名
}
```

上述代码即可取出任意十条记录的user\_id并打印出。注意分析型数据库在数据查询中是不支持SELECT \*方式查询所有列的。

## 在JAVA中连接分析型数据库

通常，在JAVA中，我们通过连接池来使用分析型数据库。在这里我们以国产的高性能连接池Druid为例来演示连接分析型数据库的方式。

```
import com.alibaba.druid.pool.*;
DruidDataSource dataSource = new DruidDataSource();
dataSource.setDriverClassName("com.mysql.jdbc.Driver");
dataSource.setUsername("my_access_key_id");
dataSource.setPassword("my_access_key_secret");
dataSource.setUrl("jdbc:mysql://mydbname-xxxx.ads-hz.aliyuncs.com:5544/my_ads_db");
// 连接数配置
dataSource.setInitialSize(5);
dataSource.setMinIdle(1);
dataSource.setMaxActive(10);
// 启用监控统计功能
dataSource.setFilters("stat");
// for mysql
dataSource.setPoolPreparedStatements(false);
// 使用心跳语句检测空闲连接
dataSource.setValidationQuery("show status like '%Service_Status%'");
dataSource.setTestWhileIdle(true);
```

如上，需要注意的是，若是在任何语言中需要使用心跳SQL来进行分析型数据库服务状态检测，请使用 `show status like '%Service_Status%'` 语句，若返回一行两列且第二列为1，则分析型数据库服务正常。

## 在Python中连接分析型数据库

在python程序中可以通过MySQLdb的module来访问AnalyticDB。关于MySQLdb的更多信息，请参考：<http://mysql-python.sourceforge.net/MySQLdb.html>

```
#!/usr/bin/python
# -*- coding: UTF-8 -*-

import MySQLdb

# 打开数据库连接
# host为目标AnalyticDB数据库实例的访问URL的域名或IP
# port为目标AnalyticDB数据库实例的访问URL的端口
# user为目标AnalyticDB数据库实例的用户账号（通常为阿里云AK的access id）
# passwd为目标AnalyticDB数据库实例的用户密码（通常为阿里云AK的access key）
# db为目标AnalyticDB数据库实例的database名称
db = MySQLdb.connect(host='test-db-89b1ac42.cn-xxx.ads.aliyuncs.com', port=9999, user='test_user',
passwd='123456', db='test_db')

# 使用cursor()方法获取操作游标
cursor = db.cursor()

# 使用execute方法执行SQL语句
cursor.execute("SELECT VERSION()")

# 使用 fetchone() 方法获取一条数据
```

```
data = cursor.fetchone()

print "Database version : %s " % data

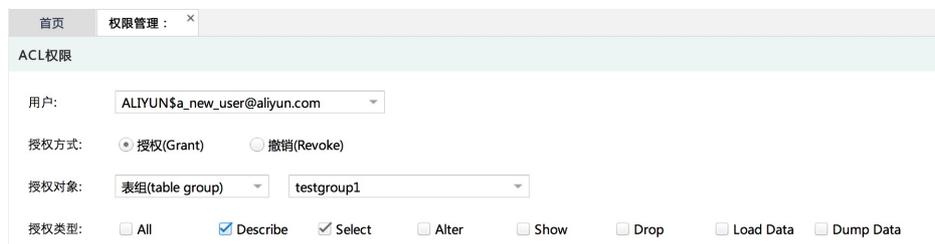
# 关闭数据库连接
db.close()
```

## 1.6 新增用户和授权

在分析型数据库中，创建一个数据库的账号为这个数据库的owner账号，拥有最大化的权限。在很多情况下，我们不希望这个账号被更多的人掌握，所以需要授权其他人仅使用较小的权限连接分析型数据库。这时我们便需要进行权限管理，完整的权限体系我们将会在第六章详细描述，这里先给出一个简单的例子。

在DMS中，左侧的对象列表中，点击用户可以看到可以访问该数据库的全部用户列表。若要新增一个授权，则我们点击菜单栏上的“权限”按钮。

在打开的窗口中，我们输入待授权的用户账号，注意由于分析型数据库将会支持多种账号来源，所以需要在输入账号时声明账号来源，例如 ALIYUN\$ 前缀代表该账号是阿里云账号。我们输入一个待授权的云账号例如 ALIYUN\$a\_new\_user@aliyun.com。然后假设我们希望这个账号对testgroup1表组下的所有表有查看元信息和只读权限，那么我们进行如下选项操作：



The screenshot shows the 'ACL权限' (ACL Permissions) management interface. It includes a breadcrumb '首页 > 权限管理' and a close button. The form contains the following fields and options:

- 用户 (User): ALIYUN\$a\_new\_user@aliyun.com
- 授权方式 (Authorization Method):  授权(Grant)  撤销(Revoke)
- 授权对象 (Authorization Object): 表组(table group) | testgroup1
- 授权类型 (Authorization Type):  All  Describe  Select  Alter  Show  Drop  Load Data  Dump Data

然后点击保存，完成授权操作。随后 a\_new\_user@aliyun.com 该用户便可使用自己的Access Key登录分析型数据库执行有权限的操作了。

## 1.7 删除表或数据库

如果您不再使用分析型数据库中的某张表，您可以执行如下SQL语句：

```
drop table dbname.tablename;
```

dbname为数据库名，tablename为要删除的表名，均为必填。一般来讲，删除表后的15分钟内，表占用的数据存储空间会释放。

如果您不想再使用分析型数据库中的某个数据库，请在删除数据库之前，先依上述步骤依次删除数据库内的所有表，之后退回到控制台的数据库选择页面，点击数据库的删除按钮，经过风控系统判断通过后，即可删除数据库。

数据库删除之后的下两个账单周期后，便不再产生此数据库的计费。